

The hPMS2 gene encodes a protein which is involved in DNA mismatch repair and is mutated in a subset of patients with hereditary nonpolyposis colon cancer (HNPCC). The previously published hPMS2 cDNA sequence lacks an upstream in-frame stop codon preceding the presumptive initiating methionine. To further evaluate the 5' terminus of the hPMS2 coding region, we isolated additional cDNA clones, RT-PCR products, and the corresponding 5' genomic segment of the hPMS2 locus. The hPMS2 gene transcripts were found to have heterogeneous but collinear 5' termini, one of which contained an in-frame termination codon preceding the initiating methionine. In addition, a gene encoding a 34.5 kDa polypeptide was found to transcriptionally initiate within hPMS2 from the opposite strand.

- 2 -

peptides from the 85 kDa protein revealed it to be the product of *hMLH1*, and this protein's molecular weight agreed with that predicted from the cDNA sequence (Bronner et.al., 1994; Papadopoulos et.al., 1994). The sequence of the peptide generated from the 110 kDa component showed it to be similar to the *hPMS2* mutL-homolog; however, the predicted molecular weight of *hPMS2* is only 95 kDa (Nicolaidis, et.al., 1994). Since the previously isolated *hPMS2* cDNA clones lacked an in-frame termination codon upstream of the presumptive initiating methionine, it was possible that the open reading frame extended further upstream. Thus there is a need in the art for further knowledge of the genetic structures of and adjacent to the known *hPMS2* gene.

SUMMARY OF THE INVENTION

It is an object of the invention to provide a novel, isolated, human gene on chromosome 7.

It is an object of the invention to provide vectors and host cells for making a novel human gene product.

It is another object of the invention to provide compositions of matter containing the human gene product.

These and other objects are provided by one or more of the embodiments described below. In one embodiment of the invention, a segment of cDNA is provided. The cDNA consists of the sequence of nucleotides shown in Figure 2.

According to another embodiment of the invention, a vector comprising the segment of cDNA which consists of the sequence of nucleotides shown in Figure 2 is provided, as well as host cells comprising the vector.

According to still another embodiment of the invention, a composition is provided. The composition consists essentially of a protein consisting of the amino acid sequence shown in Figure 2

In yet another embodiment of the invention a composition of protein *JTV1* as shown in Figure 1 is provided. The composition is free of other human proteins.

- 3 -

In another embodiment of the invention a segment of cDNA is provided which segment encodes the amino acid sequence of JTV1 protein shown in Figure 2.

cDNA probes are also provided by the present invention. The cDNA portion of said probes consists of between 15 and 1176 contiguous nucleotides of the sequence shown in SEQ ID NO:1.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the sequence of the 5' region of *hPMS2* and predicted coding region. The arrow indicates the 5' end of the previously published cDNA clone. The presumptive initiating methionine is underlined.

Figure 2 shows the sequence of *JTV1*. The sequence has been deposited in Genbank, accession number U24169. The presumptive initiating methionine is underlined.

Figure 3 demonstrates the genomic localization of *JTV1*. The genomic localization of *hPMS2* and *JTV1* were confirmed by screening somatic-cell hybrids containing various regions of human chromosome 7. Lane 1, GM10791 contains entire chromosome 7 in a chinese hamster ovary (CHO) background; lane 2, NA11440 contains 7pter>7p22 in a CHO background; lane 3, Ru-Rag4-13 contains 7cen-7pter in a murine background; lane 4, 4AF1/106/K015 contains 7cen-qter in a murine background; lane 5, GM05184.17 contains 7q21.2-qter in a CHO background; lane 6, 2068Rag22-2 contains 7q22-qter in a murine background; lane 7, human genomic DNA; lane 8, mouse genomic DNA; lane 9, CHO genomic DNA.

Figure 4 demonstrates the mapping of transcriptional start sites of *hPMS2* and *JTV1*. Sequence of the genomic region containing the 5' ends of the two genes is shown. The sequence is numbered in respect to codon 1 of *hPMS2*. Lower case letters denote intronic sequence of *JTV1* (from nt -479 to -833) and *hPMS2* (from +24 to +108). Arrows indicate the 5' ends of *hPMS2* (sense strand) and of *JTV1* (antisense strand) cDNA clones. The underlined ATG codons indicate the predicted initiating methionines for *hPMS2* (at nt +1 on the sense

- 4 -

strand) and *JTV1* (at nt -345 on the antisense strand). The sequence has been deposited in Genbank, accession number U24168.

Figure 5 shows the expression of *hPMS2* and *JTV1*. RNA from various tissues was incubated with reverse transcriptase (RT+) or in control reactions without reverse transcriptase (RT-). The cDNA was used as template for PCR with primers specific for *hPMS2* (A) and *JTV1* (B). RT-PCR products were separated by polyacrylamide gel electrophoresis.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

To investigate the upstream region from *hPMS2*, we isolated additional cDNA clones, analyzed the 5' end of *hPMS2* transcripts with PCR-based techniques, and cloned the corresponding genomic segments. In addition to clarifying the transcript, we serendipitously discovered a previously undescribed gene overlapping *hPMS2*. That gene is termed herein *JTV1*. The sequences of the *JTV1* cDNA and protein are shown in SEQ ID NOS:1 and 2, respectively.

A segment of cDNA according to the present invention refers to a contiguous stretch of deoxyribonucleotides which have a sequence as obtained upon reverse transcriptase of an RNA transcript. Such segments do not contain introns. The segment may be an isolated molecule or it can be covalently joined to other nucleic acid sequences. The segment may, for example, be replicated as part of a vector, such as a plasmid, virus, or minichromosome. The vector may be replicated within a host cell, such as a cell transformed by a recombinant DNA molecule. The host cell may be used to produce *JTV1* protein. It can also be used to study regulation of expression of *JTV1* sequences, for example by subjecting the host cell to various agents which may or may not affect the expression. Although the DNA sequence is discussed with particularity herein, it is well within the skill of the art to make small mutations, such as single nucleic acid substitutions of one of the other three nucleic acid bases, at any of the positions of the sequence. In addition, it is well within the art to make single base deletions or single base insertions, to study the effect upon protein structure and function.

If JTV1 is produced in a recombinant host cell which is not human, a composition of JTV1 protein will be produced which is free of other human proteins. If JTV1 protein is isolated from naturally producing cells, or from human host cells, then the protein can be purified, for example, using antibodies which are raised against an immunogen comprising JTV1 amino acid sequence. Any other means of purification known in the art can be used, as is desired.

DNA molecules can be made having different nucleotide sequences from that disclosed in SEQ ID NO:1, but which still encode the JTV1 protein as disclosed in SEQ ID NO:2. Using the known coding relationships between codons and amino acids and the disclosed amino acid sequence, numerous other sequences can be readily designed and produced. Such DNA molecules are within the contemplation of the subject invention.

cDNA probes can be used for hybridization studies. Typically they are labeled with a detectable marker, such as a radiolabel or a fluorescent moiety, although they need not be. The cDNA probes of the subject invention consist of at least 15 contiguous nucleotides of the sequence shown in SEQ ID NO:1. If greater specificity is desired, larger molecules of 18, 20, 25, or 30 nucleotides can be used, up to a maximum of the entire sequence of 1176 nucleotides.

JTV1 cDNAs can be used as probes to detect deletions in chromosome 7. Due to the overlapping promoter regions, large deletions of JTV1 would also be expected to affect PMS2 expression, leading to Hereditary Non-Polyposis Colorectal Cancer (HNPCC). JTV1 cDNA can be used in chromosome mapping. It can also be used to assay activity or competence of the PMS2 promoter region. The presence of JTV1 transcripts or JTV1 protein suggests that the PMS2 promoter is intact. If the PMS2 promoter is intact and PMS2 products are absent, a structural defect in the coding region is indicated.

JTV1 sequences can be used to guide homologous recombination at the PMS2 locus. For example, where a PMS2 mutation is present and therapeutic replacement with a wild-type gene is desired, PMS2 sequences can be used to provide an adjacent region of homology. Similarly, it may be desirable to target other genes to the region adjacent to PMS2. JTV1 sequences can be used to flank

such other genes, providing one or more regions of homology. If insertion of other genes is desired between the *JTV1* and the *PMS2* sequences, again, this can be accomplished using the identified sequences as homology units for homologous recombination.

Examples

Example 1

Isolation and sequence analysis of the 5' end of *hPMS2*.

Purified DNA from P1 clone 53, previously determined to contain the *hPMS2* gene (Nicolaidis, et.al., 1994), was digested with EcoRI and subcloned into the pBluescript vector (Stratagene). Clones containing the 5' region of *hPMS2* were identified by hybridization with primer A (Table 1) directed to exon 1. Restriction analysis of several positive clones showed them to be identical. The sequence of the relevant region of *hPMS2* was determined from both strands using ³⁵S α -dATP and Sequenase (USB).

Table 1. Primers used for *hPMS2*.

PRIMER NAME	STRAND	PRIMER SEQUENCE	POSITION*
A	sense	5'-cggggtgtgcatccatgg-3'	-14 - +4
B	sense	5'-gggtggagcacaacgtcg -3'	-110 - -93
C	sense	5'-ggtcacgacggagaccg-3'	-283 - -267
D	sense	5'-tgcagggtgggaagctccacacgg-3'	-414 - -392
E	sense	5'-tagctcctgccgtgcacg-3'	-448 - -431
F	sense	5'-cgctcctacctgcacgtg-3'	-487 - -470
G	antisense	5'-tagactcagtaccacctgc-3'	+90 - +107
H	sense	5'-tacagaacctgctaaggcc-3'	+24 - +42
I	antisense	5'-ttctactaactccttaccg-3'	+116 - +136
J	sense	5'-caaccatgagacacatcgc-3'	+2545 -
K	antisense	5'-aggttagtgaagactctgtc-3'	+2647 - +2666

* Relative to the presumptive initiating methionine in Figure 1.

Three clones were isolated, each containing an 8.5 kb EcoRI insert. Partial sequence analysis of one clone, pSMN, determined that it contained coding residues of *hPMS2* as well as sequences upstream of the previously designated codon 1. The presumptive initiating codon reported previously has been designated as nucleotide 1 in Figure 1. The sequence of *hPMS2* was extended 833 bp upstream of nucleotide 1. This sequence revealed an in-frame stop codon 321 nts upstream of the published initiator methionine, with no intervening methionines (Figure 1).

Example 2

Isolation of additional cDNA clones using *hPMS2* probes.

Two cDNA libraries were screened with a probe containing nt +24 to +136 of *hPMS2* generated by PCR using P1 clone 53 as template and the primers H and I (Table 1). A human small intestine random-primed cDNA library in λ GT10 (Clontech) and a HeLa oligo-dT primed cDNA library in λ ZAPII (Stratagene) were screened as described except hybridizations were carried out at 68°C and filters were washed at 65°C for 0.5 hrs (Kinzler and Vogelstein, 1989). Following plaque purification, the EcoRI inserts from the small intestine library were subcloned into pBluescript vector, while the HeLa cDNA inserts were rescued as phagemids following the manufacturer's protocol (Stratagene).

One clone was isolated from the random-primed small intestine library, and this contained nt -14 to nt +1668 of *hPMS2*. Two clones were isolated from the oligo-dT primed HeLa cDNA library. The clones began at nt -53 and ended at either nts +2722 or +2749. The HeLa cDNA library was also screened with a 430 bp probe from the 5' genomic region of *hPMS2*, containing nt -414 to +16, generated by PCR from P1 clone 53 using primers D (Table 1) and O (Table 2). The same two clones were identified, as expected. However, twelve other overlapping clones were found and appeared to represent a different transcript, named *JTVI* (Figure 2). These twelve cDNAs were approximately 1.2 kb in length and were sequenced in their entirety. All twelve ended with a polyA tract (assumed to be the 3' end) and were identical for 1.2 kb upstream. The 5' ends were located within 38 bp of each other. Comparison with *hPMS2* indicated that *JTVI* was transcribed from the opposite strand.

Table 2. Primers used for *JTV-1* cDNA amplification.

PRIMER NAME	STRAND	PRIMER SEQUENCE	POSITION*
L	sense	5'-gttctgccatgccgatg-3'	-8 - +9
M	sense	5'-ggcctttggcacgcgctac-3'	-23 - -41
N	sense	5'-accggactgcgttttcccg-3'	-111 - -129
O	sense	5'-tctcagctcgtccatgg-3'	-343 - -360
P	antisense	5'-gcagagacaggttagactc-3'	+139 - +157
Q	sense	5'-gctcctaagtgaattgccg-3'	+952 - +971
R	antisense	5'-tgacacttgacaactggcc-3'	+1068 - +1086

* Relative to the presumptive initiating methionine in Figure 2.

Example 3

JTVL

The length of one clone representative of *JTVL* (pM23NNFL) was 1233 bp and encoded an open reading frame (ORF) of 936 bp (Figure 2). The first methionine within this ORF was designated codon 1 (Figure 2) and was preceded by an in-frame termination codon 66 bp upstream. This methionine had a reasonable match to the Kozak translation initiation consensus (Kozak, 1986). The 3' end contained a polyadenylation signal (AAUAAA) starting at nucleotide 1086 followed by a polyA tail. The transcript was predicted to encode a polypeptide of 312 amino acids, with a molecular weight of 34.5 kda. Searches of nucleotide and peptide sequence databases showed that this was a novel gene, with limited homology to the glutathione S-transferase gene family.

- 10 -

Example 4

Chromosomal Mapping of *JTVI*.

The *hPMS2* locus was previously mapped to chromosome 7p22 by FISH using P1 clone 53 (Nicolaidis et.al., 1994). Because multiple *hPMS2*-related genes are located on the long arm of chromosome 7 and have conserved 5' regions (personal observation, Hori et.al., 1994), we confirmed the genomic localization of *JTVI* by PCR analysis of rodent-human somatic cell hybrid DNAs containing various regions of chromosome 7 (Scherer et.al., 1993; Powers et.al., 1993). PCR primers were chosen from the 3' untranslated region of *hPMS2* and *JTVI* and shown to amplify genomic DNA. *hPMS2* primers J and K yielded a 121 bp product and *JTVI* primers Q and R yielded a 134 bp product. PCR products for both genes were formed in those DNAs containing the 7p22 region: lines GM10791 (containing the entire human chromosome 7), NA11440 (Coriell Institute) (7p22 > 7pter) and Ru-Rag4-13 (7cen-7pter) (figure 3, lanes 1, 2, and 3). No products were observed in lines 4AF1/106/K015 (7cen-qter), GM05184.17 (7q21.2-qter), or 2068Rag22-2 (7q22-qter) (figure 3, lanes 4, 5, and 6).

Example 5

Analysis of the 5' Termini of *hPMS2* and *JTVI*.

The 5' termini of *hPMS2* transcripts were studied by standard cDNA cloning, RACE, and RT-PCR analyses. RNA was purified from tissues and cells using a guanidine isothiocyanate based method (Chomczynski and Sacchi, 1987). Reverse transcriptase-polymerase chain reaction (RT-PCR) was performed using randomly primed cDNA as template as described (Leach, et.al., 1993). RT-PCR of the 5' end of *hPMS2* was performed using a common antisense primer (I) and the sense primers (A-F) described in Table 1. RT-PCR mapping of the 5' end of *JTVI* was done using a common antisense primer P and the sense primers L-O as described in Table 2. RACE (rapid amplification of cDNA ends, Frohman, et.al., 1988) was performed on *hPMS2* using sequential antisense primers I and G (Table 1) following the manufacturer's protocol (Clontech). RACE analysis of *JTVI* was done using the antisense primer P (Table 2). Amplification products were cloned

into a T-tailed vector (InVitrogen) and sequenced using SP6 and T7 primers. Amplifications were done at 95°C for 30 sec, 56°C for 1.5 min., and 70°C for 1.5 min for 35 cycles. Reaction products were separated by electrophoresis in 6% nondenaturing polyacrylamide gels.

Figure 4 shows the sequence of the genomic region containing the transcriptional initiation sites of both *hPMS2* and *JTVI*, numbered as in Figure 1 with respect to *hPMS2*. The 5' ends of *hPMS2* cDNA clones are marked with arrowheads on the top strand. One clone began at nt -14, one at nt -24, and two at nt -53. RACE products were generated from adult brain, leukocyte, and placenta mRNA. Using an antisense primer corresponding to nt +116 to +136, multiple bands with approximately 160 to 191 bps were observed in addition to less intense bands of up to 550 bp. The sequence of four cloned RACE products demonstrated that, as expected, their 5' ends were located between nt -25 to -55. These data suggested that the majority of *hPMS2* transcripts initiated between nt -13 to -55, with a minority extending further upstream. This was confirmed by RT-PCR analysis using mRNA from HeLa cells as template. Robust RT-PCR products were amplified with sense primers whose 5' ends were at nt -14, -110, -283, and -414, (primers A, B, C, and D; Table 1) and an antisense primer corresponding to nt +90 to +107 (G). No PCR products were observed using sense primers whose 5' ends were at nt -448 or -487 (primers E and F). To ensure that primers E and F were not defective, successful amplification of genomic DNA was performed using these primers and an antisense primer (O) corresponding to nt -2 to +16.

The 5' termini of *JTVI* showed a heterogeneous pattern like that of *hPMS2*. The 5' ends of the 12 cDNA clones are indicated by arrowheads on the bottom strand in figure 4. They were located 73 to 113 nt 73 upstream of codon 1 of *JTVI*, which corresponded to nt -271 to -232 of *hPMS2*. RACE confirmed the cDNA results in that the majority of products generated using an antisense primer P corresponding to *JTVI* nt +157 were 230 to 270 bp. RT-PCR analysis was performed with antisense primer P and several sense primers (L-O) listed in Table 2. PCR products were found with sense primers whose 5' ends were at -8, -23,

and -111, (primers L, M, and N) but not with a sense primer O whose 5' end was at nt -360 with respect to *JTVI*, nt +1. The latter primer was not defective, as a genomic segment could be successfully amplified with it.

Transcripts of *hPMS2* had heterogeneous but collinear 5' termini, containing 11 to 415 nt of presumably untranslated sequence. The transcripts contained an in-frame stop codon upstream of the presumptive initiating methionines (Figure 1), making the originally described methionine the most likely translation initiator. Because no other upstream coding regions of *hPMS2* appeared to exist, the size discrepancy between that predicted from the *hPMS2* sequence and the 110 kDa *hPMS2* protein identified by Li and Modrich is likely due to post-transcriptional modifications or alternative internal exons.

Our results revealed that *hPMS2* overlaps with a novel gene, *JTVI*, transcribed from the opposite strand (Figure 4). This organization is similar to that of HUMDUG, a *mutS*-homolog found on human chromosome 5, and the dihydrofolate reductase (DHFR) gene (Fujii and Shimada, 1989). Both *hPMS2*-*JTVI* and HUMDUG-DHFR lie in a head to head arrangement, both genes are ubiquitously expressed, and both have multiple 5' termini. It has been hypothesized that DHFR and HUMDUG may be regulated via a bidirectional promoter, because a minor subset of the transcripts from the two genes overlap. The major transcripts of HUMDUG and DHFR, however, do not overlap, as is true for *hPMS2* and *JTVI*. It will be of interest to determine whether other mismatch repair genes are arranged in a head to head fashion with a contiguous gene and if *JTVI* is involved in DNA replication or repair.

Example 6

Expression of *hPMS2* and *JTVI*

The expression of *hPMS2* and *JTVI* was analyzed in a variety of mRNA samples prepared from human tissues. RT-PCR was performed on cDNA templates derived from adult brain, leukocytes, kidney, large intestine, colon, salivary gland, lung, testes and prostate using primers J and K for *hPMS2* and

primers Q and R for *JTVI* (Tables 1 and 2). Both genes were expressed in all tissues tested (Figure 5).

References

Aaltonen, L.A., Peltomaki, P., Leach, F.S., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Powell, S.M., Jen, J., Hamilton, S.R., Petersen, G.M., Kinzler, K.W., Vogelstein, B., and de la Chapelle, A. (1993). Clues to the pathogenesis of familial colorectal cancer. *Science* 260:812-816.

Chomczynski, P., and Sacchi, N. (1987). Single step method of RNA isolation by acidic guanidinium-isothiocyanate phenol-chloroform extraction. *Anal. Biochem.* 162:6-13.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoc, M.K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., Tannergard, P., Bollag, R.J., Godwin, A.R., Ward, D.C., Nordenskjold, M., Fishel, R., Kolodner, R., and Liskay, R.M. (1994). Mutation in the DNA mismatch repair gene homologue *hMLH1* is associated with hereditary non-polyposis colon cancer. *Nature* 368:258-261.

Fishel, R., Lescoc, M., Rao, M.R.S., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. *Cell* 75:1027-1038.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85:8998-9002.

- Fujii, H., and Shimada, T. (1989). Isolation and characterization of cDNA clones derived from the divergently transcribed gene in the region upstream from the human dihydrofolate reductase gene. *J. Biol. Chem.* 264:10057-10064.
- Hori, A., Han, H.-J., Sasaki, S., Shimada, M., and Nakamura, Y. (1994). Cloning, characterization and chromosomal assignment of the human genes homologous to yeast *PMS1*, a member of mismatch repair genes. *Biochem. Biophys. Res. Comm.* 204:1257-1264.
- Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 260:558-561.
- Kinzler, K.W., and Vogelstein, B. (1989). Whole genome PCR: application to the identification of sequences bound by gene regulatory proteins. *Nuc. Acid. Res.* 17:3645-3653.
- Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. *Cell* 44:283-292.
- Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., Nystrom-Lahti, M., Guan, X.-Y., Zhang, J., Meltzer, P.S., Yu, J.-W., Kao, F.-T., Chen, D.J., Cerosaletti, K.M., Fournier, R.E.K., Todd, S., Lewis, T., Leach, R.J., Naylor, S.L., Weissenbach, J., Mecklin, J.-P., Jarvinen, J.A., Petersen, G.M., Hamilton, S.R., Green, J., Jass, J., Watson, P., Lynch, H.T., Trent, J.M., de la Chapelle, A., Kinzler, K.W., and Vogelstein, B. (1993). Mutations of a *mutS* homolog in hereditary nonpolyposis colorectal cancer. *Cell* 75:1215-1225.

- Li, G.-M., and Modrich, P. (1994). Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human *mutL* homologs. *Proc. Natl. Acad. Sci. USA* 92:1950-1954.
- Lynch, H.T., Smyrk, T.C., Watson, P., Lanspa, S.J., Lynch, J.F., Cavalieri, R.J., and Boland, C.R. (1993). Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: An updated review. *Gastroenterology* 104:1535-1549.
- Modrich, P. (1995). Mismatch repair, genetic stability, and cancer. *Science* 266:1959-1960.
- Nicolaides, N.C., Papadopoulos, N., Liu, B., Wei, Y.-F., Carter, K.C., Ruben, S.M., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., Venter, J.C., Dunlop, M.G., Hamilton, S.R., Petersen, G.M., de la Chapelle, A., Vogelstein, B., and Kinzler, K.W. (1994). Mutations of two *PMS* homologues in hereditary nonpolyposis colon cancer. *Nature* 371:75-80.
- Palombo, F., Hughes, M., Jiricny, J., Truong, O., and Hsuan, J. (1994). Mismatch repair and cancer. *Nature* 367:417-418.
- Papadopoulos, N., Nicolaides, N.C., Wei, Y.-F., Ruben, S.M., Carter, K.C., Rosen, W.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., Venter, J.C., Hamilton, S.R., Petersen, G.M., Watson, P., Lynch, H.T., Peltomaki, P., Mecklin, J.-P., de la Chapelle, A., Kinzler, K.W., and Vogelstein, B. (1994). Mutation of a *mutL* homolog in hereditary colon cancer. *Science* 263:1625-1629.
- Parsons, R., Li, G.-M., Longley, M.J., Fang, W.-H., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K.W., Vogelstein, B., and Modrich, P. (1993).

Hypermutable and mismatch repair deficiency in RER+ tumor cells. *Cell* 75: 1227-1236.

Powers, P.A., Scherer, S.W., Tsui, L.-C., Gregg, R.G., Hogan, K. (1994). Localization of the gene encoding the α_1/δ subunit (CACNL2A) of the human skeletal muscle voltage-dependent Ca^{2+} channel to chromosome 7q21-22 by somatic cell hybrid analysis. *Genomics* 19:192-193.

Scherer, S.W., Neufeld, E.J., Lievens, P.M.-J., Orkin, S.H., Kim, J., and Tsui, L.-C. (1993). Regional localization of the CCAAT displacement protein gene (CUTL1) to 7q22 by analysis of somatic cell hybrids. *Genomics* 15:695-696.

Shibata, D., Peinado, M.A., Ionov, Y., Malkhosyan, S., and Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nature Genet.* 6:273-281.

Thibodeau, S.N., Bren, G., and Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* 260:816-819.

Umar, A., Boyer, J.C., Thomas, D.C., Nguyen, D.C., Risinger, J.I., Boyd, J., Ionov, Y., Perucho, M., and Kunkel, T.A. (1994). Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability. *J. Biol. Chem.* 269:14367-14370.

-17-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Vogelstein, Bert
Kinzler W., Kenneth
Nicolaides C., Nicholas
- (ii) TITLE OF INVENTION: Human JTV1 Gene Overlaps PMS2 Gene
- (iii) NUMBER OF SEQUENCES: 5
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Banner & Allegretti, LTD.
 - (B) STREET: 1001 G Street, NW
 - (C) CITY: Washington DC
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 20001
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Kagan A., Sarah
 - (B) REGISTRATION NUMBER: 32,141
 - (C) REFERENCE/DOCKET NUMBER: 1107.49697
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 202-508-9100
 - (B) TELEFAX: 202-508-9299

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 384 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 46..384

-18-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTACCTGGTA CATCGGCATG GCAGAACCRA AGCAAAAGGG GGTAG CGC GTG CCA	54
Arg Val Pro	
1	
AAG GCC AAC GCT CAG AAA CCG TCA GAG GTC ACG ACC GAG ACC GGC CAC	102
Lys Ala Asn Ala Gln Lys Pro Ser Glu Val Thr Thr Glu Thr Gly His	
5 10 15	
CTC CCT TCT GAC CCT GCT GCG GGC GTT CCG GAA AAC CCA GTC CCG TGT	150
Leu Pro Ser Asp Pro Ala Ala Gly Val Arg Glu Asn Ala Val Arg Cys	
20 25 30 35	
GCT CTG ATT GGC CCA GGC TCT TTG ACG TCA CGA AGT CGA CCT TTG ACA	198
Ala Leu Ile Gly Pro Gly Ser Leu Thr Ser Arg Ser Arg Pro Leu Thr	
40 45 50	
GAG CCA ATA GGC GAA AAG GAG AGA CCG GAA GTA TTT TTG CCG CCC CCG	246
Glu Pro Ile Gly Glu Lys Glu Arg Arg Glu Val Phe Leu Pro Pro Arg	
55 60 65	
CCG GAA AGG GTG GAG CAC AAC GTC GAA AGC AGC CAA TGG GAG TTC AGG	294
Pro Glu Arg Val Glu His Asn Val Glu Ser Ser Gln Trp Glu Phe Arg	
70 75 80	
AGG CCG AGC GCC TGT GGG AGC CCT GGA GGG AAC TTT CCC AGT CCC CGA	342
Arg Arg Ser Ala Cys Gly Ser Pro Gly Gly Asn Phe Pro Ser Pro Arg	
85 90 95	
GGC GGA TCG GGT GTT GCA TCC ATG GAG CGA GCT GAG AGC TCG	384
Gly Gly Ser Gly Val Ala Ser Met Glu Arg Ala Glu Ser Ser	
100 105 110	

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 113 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Arg Val Pro Lys Ala Asn Ala Gln Lys Pro Ser Glu Val Thr Thr Glu	
1 5 10 15	
Thr Gly His Leu Pro Ser Asp Pro Ala Ala Gly Val Arg Glu Asn Ala	
20 25 30	
Val Arg Cys Ala Leu Ile Gly Pro Gly Ser Leu Thr Ser Arg Ser Arg	
35 40 45	
Pro Leu Thr Glu Pro Ile Gly Glu Lys Glu Arg Arg Glu Val Phe Leu	
50 55 60	
Pro Pro Arg Pro Glu Arg Val Glu His Asn Val Glu Ser Ser Gln Trp	
65 70 75 80	
Glu Phe Arg Arg Arg Ser Ala Cys Gly Ser Pro Gly Gly Asn Phe Pr	
85 90 95	

-19-

Ser Pro Arg Gly Gly Ser Gly Val Ala Ser Met Glu Arg Ala Glu Ser
 100 105 110

Ser

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1233 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 114..1049

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCGAACGCCC GCAGCAGGGT CAGAAGGGAG GTGGCCGGTC TCCCTCGTGA CCTCTGACGG	60
TTTCTGAGCG TTGGCCTTTG GCACGCGCTA CACCCTTTTG CTTTGGTTCT GCC ATG	116
	Met
	1
CCG ATG TAC CAG GTA AAG CCC TAT CAC GGG GGC GGC GCG CCT CTC CGT	164
Pro Met Tyr Gln Val Lys Pro Tyr His Gly Gly Gly Ala Pro Leu Arg	
	5 10 15
GTG GAG CTT CCC ACC TGC ATG TAC CGG CTC CCC AAC GTG CAC GGC AGG	212
Val Glu Leu Pro Thr Cys Met Tyr Arg Leu Pro Asn Val His Gly Arg	
	20 25 30
AGC TAC GGC CCA GCG CCG GGC GCT GGC CAC GTG CAG GAA GAG TCT AAC	260
Ser Tyr Gly Pro Ala Pro Gly Ala Gly His Val Gln Glu Glu Ser Asn	
	35 40 45
CTG TCT CTG CAA GCT CTT GAG TCC CGC CAA GAT GAT ATT TTA AAA CGT	308
Leu Ser Leu Gln Ala Leu Glu Ser Arg Gln Asp Asp Ile Leu Lys Arg	
	50 55 60 65
CTG TAT GAG TTG AAA GCT GCA GTT GAT GGC CTC TCC AAG ATG ATT CAA	356
Leu Tyr Glu Leu Lys Ala Ala Val Asp Gly Leu Ser Lys Met Ile Gln	
	70 75 80
ACA CCA GAT GCA GAC TTG GAT GTA ACC AAC ATA ATC CAA GCG GAT GAG	404
Thr Pro Asp Ala Asp Leu Asp Val Thr Asn Ile Ile Gln Ala Asp Glu	
	85 90 95
CCC ACG ACT TTA ACC ACC AAT GCG CTG GAC TTG AAT TCA GTG CTT GGG	452
Pro Thr Thr Leu Thr Thr Asn Ala Leu Asp Leu Asn Ser Val Leu Gly	
	100 105 110

-20-

AAG GAT TAC GCG GCG CTG AAA GAC ATC GTG ATC AAC GCA AAC CCG GCC Lys Asp Tyr Gly Ala Leu Lys Asp Ile Val Ile Asn Ala Asn Pro Ala 115 120 125	500
TCC CCT CCC CTC TCC CTG CTT GTG CTG CAC AGG CTG CTC TGT GAG CAC Ser Pro Pro Leu Ser Leu Leu Val Leu His Arg Leu Leu Cys Glu His 130 135 140 145	548
TTC AGG GTC CTC TCC ACG GTG CAC ACG CAC TCC TCG GTC AAG AGC GTG Phe Arg Val Leu Ser Thr Val His Thr His Ser Ser Val Lys Ser Val 150 155 160	596
CCT GAA AAC CTT CTC AAG TGC TTT GGA GAA CAG AAT AAA AAA CAG CCC Pro Glu Asn Leu Leu Lys Cys Phe Gly Glu Gln Asn Lys Lys Gln Pro 165 170 175	644
CGC CAA GAC TAT CAG CTC GGA TTC ACT TTA ATT TGG AAG AAT GTG CCG Arg Gln Asp Tyr Gln Leu Gly Phe Thr Leu Ile Trp Lys Asn Val Pro 180 185 190	692
AAG ACG CAG ATG AAA TTC AGC ATC CAG ACG ATG TGC CCC ATC GAA GGC Lys Thr Gln Met Lys Phe Ser Ile Gln Thr Met Cys Pro Ile Glu Gly 195 200 205	740
GAA GGG AAC ATT GCA CGT TTC TTG TTC TCT CTG TTT GGC CAG AAG CAT Glu Gly Asn Ile Ala Arg Phe Leu Phe Ser Leu Phe Gly Gln Lys His 210 215 220 225	788
AAT GCT GTC AAC GCA ACC CTT ATA GAT AGC TGG GTA GAT ATT GCG ATT Asn Ala Val Asn Ala Thr Leu Ile Asp Ser Trp Val Asp Ile Ala Ile 230 235 240	836
TTT CAG TTA AAA GAG GGA AGC AGT AAA GAA AAA GCC GCT GTT TTC CGC Phe Gln Leu Lys Glu Gly Ser Ser Lys Glu Lys Ala Ala Val Phe Arg 245 250 255	884
TCC ATG AAC TCT GCT CTT GGG AAG AGC CCT TGG CTC GCT GCG AAT GAA Ser Met Asn Ser Ala Leu Gly Lys Ser Pro Trp Leu Ala Gly Asn Glu 260 265 270	932
CTC ACC GTA GCA GAC GTG GTG CTG TGG TCT GTA CTC CAG CAG ATC GGA Leu Thr Val Ala Asp Val Val Leu Trp Ser Val Leu Gln Gln Ile Gly 275 280 285	980
GGC TGC AGT GTG ACA GTG CCA GCC AAT GTG CAG AGG TGG ATG AGG TCT Gly Cys Ser Val Thr Val Pro Ala Asn Val Gln Arg Trp Met Arg Ser 290 295 300 305	1028
TGT GAA AAC CTG GCT CCT TTT TAACACGGCC CTCAGCTCC TTAAGTGAAT Cys Glu Asn Leu Ala Pro Phe 310	1079
TGCCGTAAC TATTTTAAAG GGTTTAGATT TTAAGAATGG TGCTCTTTCA TGCCTATTAT	1139
CAGTAAGGGG ACTTGTATTA GAGTCAGAGT CTTTATTATT AGGCCAGTTC TCAAGTGTCA	1199
ATAAAGCGC ATCATGTAAT TTAAAAAAA AAAA	1233

(2) INFORMATION FOR SEQ ID NO:4:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312 amino acids

-21-

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Pro Met Tyr Gln Val Lys Pro Tyr His Gly Gly Gly Ala Pro Leu
 1           5           10           15
Arg Val Glu Leu Pro Thr Cys Met Tyr Arg Leu Pro Asn Val His Gly
      20           25           30
Arg Ser Tyr Gly Pro Ala Pro Gly Ala Gly His Val Gln Glu Glu Ser
      35           40           45
Asn Leu Ser Leu Gln Ala Leu Glu Ser Arg Gln Asp Asp Ile Leu Lys
      50           55           60
Arg Leu Tyr Glu Leu Lys Ala Ala Val Asp Gly Leu Ser Lys Met Ile
      65           70           75           80
Gln Thr Pro Asp Ala Asp Leu Asp Val Thr Asn Ile Ile Gln Ala Asp
      85           90           95
Glu Pro Thr Thr Leu Thr Thr Asn Ala Leu Asp Leu Asn Ser Val Leu
      100          105          110
Gly Lys Asp Tyr Gly Ala Leu Lys Asp Ile Val Ile Asn Ala Asn Pro
      115          120          125
Ala Ser Pro Pro Leu Ser Leu Leu Val Leu His Arg Leu Leu Cys Glu
      130          135          140
His Phe Arg Val Leu Ser Thr Val His Thr His Ser Ser Val Lys Ser
      145          150          155          160
Val Pro Glu Asn Leu Leu Lys Cys Phe Gly Glu Gln Asn Lys Lys Gln
      165          170          175
Pro Arg Gln Asp Tyr Gln Leu Gly Phe Thr Leu Ile Trp Lys Asn Val
      180          185          190
Pro Lys Thr Gln Met Lys Phe Ser Ile Gln Thr Met Cys Pro Ile Glu
      195          200          205
Gly Glu Gly Asp Ile Ala Arg Phe Leu Phe Ser Leu Phe Gly Gln Lys
      210          215          220
His Asn Ala Val Asn Ala Thr Leu Ile Asp Ser Trp Val Asp Ile Ala
      225          230          235          240
Ile Phe Gln Leu Lys Glu Gly Ser Ser Lys Glu Lys Ala Ala Val Phe
      245          250          255
Arg Ser Met Asn Ser Ala Leu Gly Lys Ser Pro Trp Leu Ala Gly Asn
      260          265          270
Glu Leu Thr Val Ala Asp Val Val Leu Trp Ser Val Leu Gln Gln Ile
      275          280          285

```

-22-

Gly Gly Cys Ser Val Thr Val Pro Ala Asn Val Gln Arg Trp Met Arg
 290 295 300

Ser Cys Glu Asn Leu Ala Pro Phe
 305 310

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: mRNA

(B) LOCATION: complement (1..900)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACACCCGGCC AATTCTGTA TTTTAGTAG AGACGAGGTT TTACCATGTT GGCCAGGCTA	60
GTCTCGAACT CCTGACCTCA GGTGATCCGC CCGCCTCGGC CTCCCAAAGT GCTGGGATTA	120
CAGGCGTGAG CCACGGCGCC CGGCCTGGAT AAATCTTTA AAAGATAAAA GTCTGAGTGA	180
GTCCCTGGCC GCGCGGCACA GATGCCGGGG TGGGGCCGTG AACCGGTTGG GACCGGCTCG	240
CTCCGGCCCTG GGGGGACCCG GGGCAGCAGC CGGTCCCGGC GCGTCCGCAC TGGGCGGGGG	300
GCCCCGGCGT CCTACCTGCA CGTGGCCAGG CCGGGCGCTG GGCCGTAGCT CCTGCCGTGC	360
ACGTTGGGGA GCGGTACAT GCAGGTGGGA AGCTCCACAC GGAGAGGCGC GCGCCCCCGG	420
TGATAGGGCT TTACCTGGTA CATCGGCATG GCAGAACCAA AGCAAAAGGG GGTAGCGCGT	480
GCCAAAGGCC AACGCTCAGA AACCGTCAGA GGTACGAGC GAGACCGGCC ACCTCCCTTC	540
TGACCCTGCT GCGGGCGTTC GGGAAAACGC AGTCCGGTGT GCTCTGATTG GCCCAGGCCC	600
TTTGACGTCA CGAAGTCGAC CTTTGACAGA GCCAATAGGC GAAAAGGAGA GACGGGAAGT	660
ATTTTTGCCC CCGCGCCCGG AAAGGGTGGA GCACAACGTC GAAAGCAGCC AATGGGAGTT	720
CAGGAGGCGC AGCGCCTGTG GGAGCCCTCG AGGGAACCTT CCCACTCCCC GAGGCGGATC	780
GGGTGTTGCA TCCATCGAGC GAGCTGAGAG CTCGAGGTGA GCGGGGCTCG CAGTCTTCCG	840
GTGTCCCTTC TCGCGCGCCC TCTTTGAGAC CCACGGCATT CCAACCTCCC TGGAAATGGG	900

CLAIMS

1. A segment of cDNA consisting of the nucleotide sequence shown in Figure 2.
2. A vector comprising the segment of DNA of claim 1.
3. A host cell which comprises the vector of claim 2.
4. A composition consisting essentially of a protein consisting of the amino acid sequence shown in Figure 2.
5. A composition of protein *JTVI* as shown in Figure 1, wherein said composition is free of other human proteins.
6. A segment of cDNA which encodes the amino acid sequence of JTV1 protein shown in Figure 2.
7. A cDNA probe wherein said cDNA consists of between 15 and 1176 contiguous nucleotides of the sequence shown in SEQ ID NO:1.

-370 TTA CCT GGT ACA TCG GCA TGG CAG AAC CAA AGC AAA AGG GGG TAG CGC * R --322
 V P K A N A Q K P S E V T T E T
 -321 GTG CCA AAG GCC AAC GCT CAG AAA CCG TCA GAG GTC ACG ACG GAG ACC --273
 G H L P S D P A A G V R E N A V
 -272 GGC CAC CTC CCT TCT GAC CCG GGC GTT CGG GAA AAC GCA GTC --224
 R C A L I G P G S L T S R S R P
 -223 CGG TGT GCT CTG ATT GGC CCA GGC TCT TTG ACG TCA CGA AGT CGA CCT --175
 L T E P I G E K E R R E V F L P
 -174 TTG ACA GAG CCA ATA GGC GAA AAG GAG AGA CGG GAA GTA TTT TTG CCG --126
 P R P E R V E H N V E S S Q W E
 -125 CCC CGC CCG GAA AGG GTG GAG CAC AAC GTC GAA AGC AAG CAA TGG GAG --77
 F R R R S A C G S P G G N F P S
 -76 TTC AGG AGG CGG AGC GCC TGT GGT GGG AGC CCT GGA GGG AAC TTT CCC AGT --28
 P R G G S G V A S M E R A E S S
 -27 CCC CGA GGC GGA TCG GGT GTT GCA TCC ATG GAG CCA GCT GAG AGC TCG +21

1 / 5

Figure 1

2 / 5

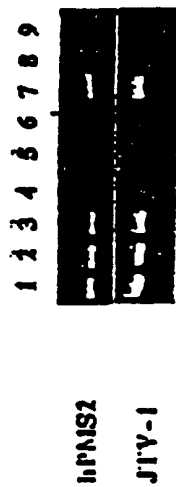
```

-113 CC GAA CGC CGC CAG CAG GGT CAG AAG GGA GGT GGC CGC TCT CCG TCG TGA CCG CTC ACG GTT TCT GAG CCG TGG CCG TTT GCA -31
1  CGC GCT ACA CCC TTT TGC TTT GGT TCT GCT GCT CGC ATG TAG CAG GTA AAG CCG TAT CAC GCG GGC GGC CCG CTC CTC CCG GTG 18
+54 E L P T C M Y R L P M V M G R S Y G P A P G A G M V Q E 46
+55 CAG CTT CCC ACC TGC ATG TAC CCG CTC CCC AAC GTC CAC GGC AGG AEC TAC GGC CCA GCG CCG GGC GGT GGC CAC GTG CAG GAA +130
47 E S M L S L Q A L E S R G D D I L K R L Y E L K A A V D 74
+139 GAG TCT AAC CTC TCT CTC CAA GCT CTT GAG TCC CCG CAA GAT GAT ATT TTA AAG CCG CTC TAT CAG TTA AAG GCT GCA GTT CAT +222
75 G L S K M I Q T P D A D L D V T M I I Q A D E P T T L T 102
+223 GGC CTC TCC AAG ATG ATT CAA ACA CCA GAT GCA GAT TTA ACC AAC ATA ATC CAA GCG GAT GAG CCG CCG ACT TTA ACC +306
103 T M A L D L N S V L G K D Y G A L K D I V I N A M P A S 130
+307 ACC AAT GCG CTC GAC TTA AAT TCA GTG CTT GCG AAG GAT TAC GCG GCG CTC AAG AAC ATC GTG ATC AAC GCA AAC CCG GCG TCC +390
131 P L S L L L V L N R L L C E H F R V L S T V M T M S S V 158
+391 CCG CTC TCC CTC CTT GTC CAC AAG CTC CTC TGT GAG CAC TTC AAG GTG TCC ACG GTG CAC ACG CAC TCC TCG GTC +474
159 K S V P E M L L K C F Q E Q M K K Q P R Q D Y Q L Q F T 186
+475 AAG AGC GTG CCG GAA AAC CTT CTC AAG TGC TTT GGA GAA CAG AAT AAA AAG CAG CCC CCG CCA CAC TAT CAG CTC GGA TTC ACT +558
187 L I M K M V P K T Q M K F S I Q T M C P I E Q E Q M I A 214
+559 TTA ATT TGG AAG AAT GTG CCG AAG ACG CAG ATG AAG TTC AGC ATC CAG ACG ATG TGC CCG ATC GAA GGC GAA GCG AAC ATT GCA +642
215 R F L F S L F G Q K M N A V N A T L I D S W V D I A I F 242
+643 CCG TTC TTC TCT TCT CCG CAG AAG CAG AAT GCT GTC AAC GCA ACC CTT ATA GAT AGC TGG GTA GAT ATT GCG ATT TTT +726
243 Q L K E G S S K A A V F R S M S A L G K S P U L A 270
+727 CAG TTA AAG GGA AGC AGT AAA GAA AAC GCG GCT GTT TTC CCG TCC ATG AAC TGT GCT CTT CCG AAG AGC CCG CTC TCG CTC GCT +810
271 G M E L T V A D V V L W S Y L Q Q I G G C S V T V P A M 298
+811 GCG AAT GAA CTC ACC GTA CCA GAC GTG GTG TCG TGT GTA CTC CAG CAG ATC GAA GGC TGC AGT GTG ACA GTG CCA GCG AAT +896
299 V Q R W M R S C E M L A P F 312
+895 GCG AAG TGG ATG AGG TCT TGT GAA AAC CTC GCT CCG TTT TAA CAC GCG CCG CAA GCT CCG TAA GTG AAT TCG CCG AAC TGA +978
+979 TTT TAA AGG GTT TAG ATT TTA AGA ATG GTG CTC TTT CAG TAT TAT CAG TAA GCG GAC TTA TAT TAG AGT CAG AGT CTT TTT +1063
+1063 ATT TAA GCG AGT TGT CAA GTG TCA ATA AAG GCG CAT CAT GTA ATT TAA AAG AAA AAA A  +1120

```

Figure 2

Figure 3



4 / 5

[illegible]

Figure 4

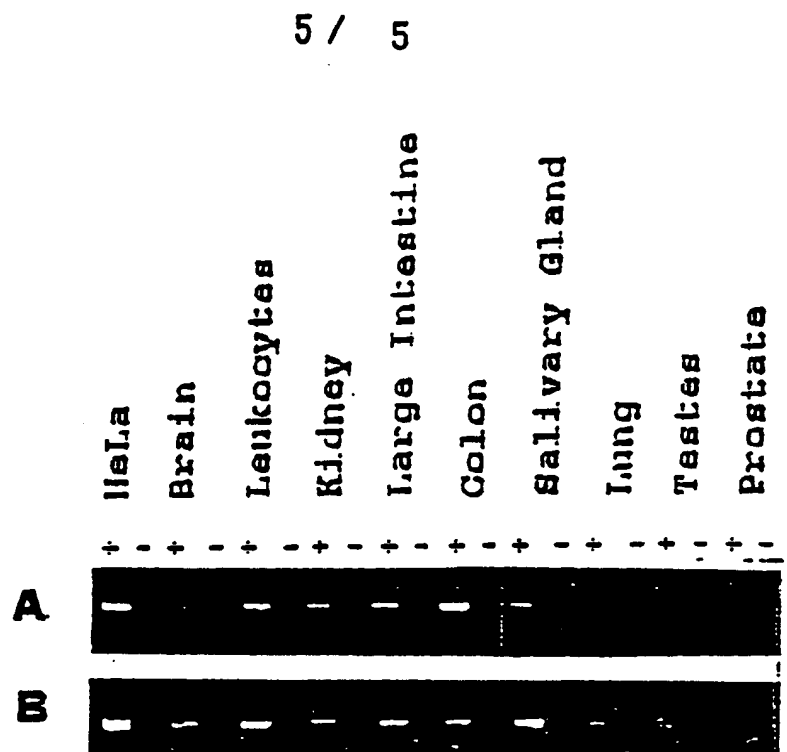


Figure 5

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/13598

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 C12N1/21 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	GENOMICS, vol. 29, 20 September 1995, pages 329-334, XP000615435 NICOLAIDES N.C. ET AL.: "Analysis of the 5' region of PMS2 reveals heterogeneous transcripts and a novel overlapping gene." see the whole document	1-7
X	--- EMBL Database entry HS321180 Accession number R84321; 16 August 1992 HILLIER ET AL.: 'The WashU-Merck EST Project.' XP002021622 see nucleotide sequence --- -/-	7

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

Z document member of the same patent family

Date of the actual completion of the international search

19 December 1996

Date of mailing of the international search report

06.01.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Mandl, B

INTERNATIONAL SEARCH REPORT

International Application No.

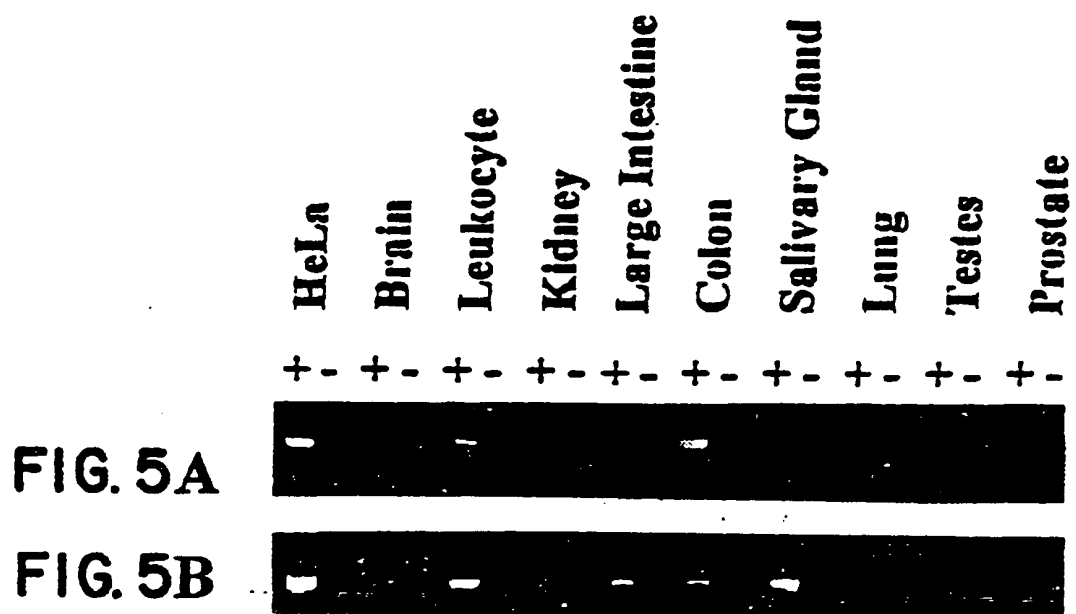
PCT/US 96/13598

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EMBL Database entry HS461263 Accession number N26461 HILLIER L. ET AL.: 'The WashU-Merck EST project.' XP002021623 see nucleotide sequence ---	7
A	NATURE, vol. 371, 1 September 1994, pages 75-80, XP002021621 NICOLAIDES ET AL.: 'Mutations of two PMS homologues in hereditary nonpolyposis colon cancer.' cited in the application see the whole document -----	1-7

+48 gtgtccctctcgcgccctcttgagaccacggcatccaaacctccctggaaatggg 3
cacagggagagcgcgcggaactctgggtccgtaaagtggagggacctttaccc 5

7 / 7



INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 96/13598

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/47 C12N1/21 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	GENOMICS, vol. 29, 20 September 1995, pages 329-334, XP000615435 NICOLAIDES N.C. ET AL.: "Analysis of the 5' region of PMS2 reveals heterogeneous transcripts and a novel overlapping gene." see the whole document ---	1-7
X	EMBL Database entry HS321180 Accession number R84321; 16 August 1992 HILLIER ET AL.: "The WashU-Merck EST Project." XP002021622 see nucleotide sequence --- -/--	7

☒ Further documents are listed in the continuation of box C.☐ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

19 December 1996

Date of mailing of the international search report

06.01.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 1
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 cpu nl.
Fax (+ 31-70) 340-3016

Authorized officer

Mandi, B

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 96/13598

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EMBL Database entry HS461263 Accession number N26461 HILLIER L. ET AL.: 'The WashU-Merck EST project.' XP002021623 see nucleotide sequence ---	7
A	NATURE, vol. 371, 1 September 1994, pages 75-80, XP002021621 NICOLAIDES ET AL.: "Mutations of two PMS homologues in hereditary nonpolyposis colon cancer." cited in the application see the whole document -----	1-7